



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Heavy metals in submicronic particulate matter (PM₁) from a Chinese metropolitan city predicted by machine learning models

Huiming Li ^a, Qian'ying Dai ^b, Meng Yang ^{c,d}, Fengying Li ^{c,d}, Xuemei Liu ^b, Mengfan Zhou ^b, Xin Qian ^{b,c,*}

^a School of Environment, Nanjing Normal University, Nanjing, 210023, China

^b State Key Laboratory of Pollution Control and Resources Reuse, School of the Environment, Nanjing University, Nanjing, 210023, China

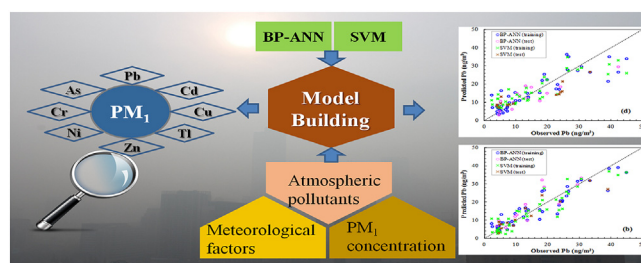
^c Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology (CICAET), Nanjing University of Information Science & Technology, Nanjing, 210044, China

^d Jiangsu Key Laboratory of Atmospheric Environment Monitoring and Pollution Control, School of Environmental Science and Engineering, Nanjing University of Information Science & Technology, Nanjing, 210044, China

HIGHLIGHTS

- No noncarcinogenic or carcinogenic risk resulted from inhalation of metals in PM₁.
- Many metals had better simulation effects when atmospheric pollutants as inputs.
- BP-ANN and SVM models both performed better for Pb, Tl and Zn than for Ti and V.
- Predicted metal contents were lower during COVID-19 outbreak than in 2018 or 2019.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 2 May 2020

Received in revised form

28 June 2020

Accepted 29 June 2020

Available online 8 July 2020

Handling Editor: R Ebinghaus

Keywords:

Airborne particle-bound metals

Back propagation artificial neural network

Support vector machine

Simulation

Health risk

ABSTRACT

The aim of this study was to establish a method for predicting heavy metal concentrations in PM₁ (aerosol particles with an aerodynamic diameter $\leq 1.0 \mu\text{m}$) based on back propagation artificial neural network (BP-ANN) and support vector machine (SVM) methods. The annual average PM₁ concentration was $26.31 \mu\text{g}/\text{m}^3$ (range: $7.00\text{--}73.40 \mu\text{g}/\text{m}^3$). The concentrations of most metals were higher in winter and lower in autumn and summer. Mn and Ni had the highest noncarcinogenic risk, and Cr the highest carcinogenic risk. The hazard index was below safe limit, and the integrated carcinogenic risk was less than precautionary value. There were no obvious differences in the simulation performances of BP-ANN and SVM models. However, in both models many elements had better simulation effects when input variables were atmospheric pollutants (SO_2 , NO_2 , CO , O_3 and $\text{PM}_{2.5}$) rather than PM₁ and meteorological factors (temperature, relative humidity, atmospheric pressure and wind speed). Models performed better for Pb, Tl and Zn, as evidenced by training R and test R values consistently >0.85 , whereas their performances for Ti and V were relatively poor. Predicted results by the fully trained models showed atmospheric heavy metal pollution was heavier in December and January and lighter in August and July of 2019. For the period covering the COVID-19 outbreak in China, from January to March 2020, most of the predicted element concentrations were lower than in 2018 and 2019, and the concentrations of nearly all

* Corresponding author. State Key Laboratory of Pollution Control and Resources Reuse, School of the Environment, Nanjing University, Nanjing, 210023, PR China.
E-mail address: xqian@nju.edu.cn (X. Qian).

metals were lowest during the nationwide implementation of countermeasures taken against the pandemic.

© 2020 Elsevier Ltd. All rights reserved.

1. Introduction

Urban air pollution is one of the most serious environmental issues confronting humankind in the face of rapid economic development (Fuentes et al., 2020). In 2016, particulate matter (PM) in ambient air was classified as a group 1 contaminant by the World Health Organization (WHO) and the International Agency for Research on Cancer (IARC) (Feng et al., 2016). The actual health effects of PM depend on its chemical composition and the ability of the respiratory tract to remove inhaled PM, which depends in turn on its size (Fang et al., 2017; Kelly and Fussell, 2012). Compared with fine particulate matter (PM_{2.5}, particles having an aerodynamic diameter $\leq 2.5 \mu\text{m}$), submicronic particulate matter (PM₁, particles with an aerodynamic diameter $< 1 \mu\text{m}$) may travel deeper into the lungs, reaching lower bronchial and alveolar regions (Oberdörster et al., 2005). Moreover, the greater surface area per mass of PM₁ allows it to readily carry potentially toxic compounds, including those of anthropogenic origin, thus posing a serious health risk to exposed populations (Caggiano et al., 2019). However, despite its greater health hazard, PM₁ is not considered by air quality standards and therefore not measured routinely. Consequently, while detailed information on potentially toxic constituents in PM₁, such as heavy metals, is vital to assess its adverse health effects, the available data are scarce (Caggiano et al., 2010, 2019; Sarti et al., 2015).

The traditional geochemical methods used in heavy metal determinations are relatively complex, time-consuming and expensive such that simpler, more cost-efficient approaches are needed. In recent years, machine learning models have been applied to predict the concentrations of airborne-pollutants such as SO₂ (Li et al., 2020), NO₂ (Yeganeh et al., 2018), O₃ (Lu and Wang, 2014), CO (Yeganeh et al., 2012), PM_{2.5} (Niu et al., 2016), airborne heavy metals (Li et al., 2017) and atmospheric nitrogen (Palani et al., 2011) as well as the air quality index (Wu and Lin, 2019), because the necessary input variables are more easily accessed, in contrast to numerical forecast models (Li et al., 2017). However, few studies have focused on the potential use of easily obtained atmospheric pollutants and meteorological factors to characterize atmospheric heavy metals statistically (Leng et al., 2017). Among the machine learning models, artificial neural network (ANN) is an adaptive system that uses neuron nodes to associate memory, nonlinear mapping, classified recognition and optimization design. One of the most widely used ANN methods is back-propagation (BP), and the flexibility of BP-ANN has been exploited to identify complex nonlinear relationships in observation-based data (Luna et al., 2014). However, this approach also suffers from problems inherent to its architecture, such as overfitting the training data, a lack of network optimization and poor generalizability (Ye et al., 2020). The support vector machine (SVM) algorithm developed by the machine learning community (Cortes and Vapnik, 1995) achieves global optimization and avoids the overfitting that limits the ANN framework (Taghvaei et al., 2016). Nonetheless, to date, neither of these statistical models has been applied to predict atmospheric heavy metals in submicronic particles.

Therefore, the main objectives of this study were: (1) to investigate the pollution levels and health risk of heavy metals in PM₁ from Nanjing, China; (2) to develop methods of simulating particle-

bound elements using BP-ANN and SVM, with atmospheric pollutants and meteorological factors as input variables and (3) to predict the temporal variation of heavy metals in PM₁ by using appropriate fully trained models.

2. Materials and methods

2.1. Sample collection

Nanjing is a megacity with a population of more than 8.1 million. It is a comprehensive industrial production center and the main transportation hub in southeast China. There are five main industries in Nanjing, including electronics, petrochemical, steel, automobile and electric power. The climate is north subtropical monsoon; the average annual temperature is 16 °C and the average annual precipitation 1106 mm. The sampling site for this study was Nanjing University (32°07'08.59" N, 118°56'50.61" E) (Leng et al., 2017; Li et al., 2017), located near the northern industrial zones and thus close to several major industrial point sources. The university's Xianlin Campus is surrounded by residential neighborhoods but is located < 6.0 km from the Nanjing Jinling Petrochemical Engineering Company. The campus is ~ 0.2 km away from the freeway where the average daily traffic volumes exceed 5000.

PM₁ samples were collected on quartz microfiber filters using a low-volume PM₁ sampler (model 600/DGDP0101K, AMS Technologies Co., Ltd., Italy) at a flow rate of 38 L/min. In order to ensure the sampler works normally and the PM mass is enough for analysis, continuous sampling was conducted for a total of 72 h during winter, 24 h per day; 96 h during spring, 24 h per day; and 120 h during summer and autumn, 24 h per day, respectively. PM₁ samples were collected over a 1-year period, from December 6, 2017 to February 6, 2018 (winter), from March 12 to May 25, 2018 (spring), from June 5 to August 29, 2018 (summer) and from September 1 to November 24, 2018 (autumn). To ensure that the samples were representative, sampling was not conducted during the wet and windy weather that can occur in the study area. A total of 57 p.m.₁ samples were collected. Each sample filter was conditioned for 48 h in a desiccator at 25 °C and 40% relative humidity before and after sampling, then weighed to determine the mass of PM₁. Hourly atmospheric pollutant concentrations (NO₂: 17i, Thermo-Fisher, USA; SO₂: 450i, Thermo-Fisher, USA; O₃: 49i, Thermo-Fisher, USA; PM_{2.5}: 5014i, Thermo-Fisher, USA) and meteorological data (wind speed: 03002-L20, R.M Young, USA; temperature and relative humidity: HMP155A, Finland; atmospheric pressure: CS106, USA) were recorded simultaneously at an automatic air quality monitoring station located about 2 m from the sampling site.

2.2. Heavy metal analysis

The PM samples were digested with a mixture of HNO₃, HCl and HF to release the metal elements. Zn concentrations were measured using inductively coupled plasma atomic emission spectrometry (PerkinElmer SCIEX, Optima 5300 DV, Norway), with a minimum detection limit of 0.001 mg/L. As, Cd, Co, Cr, Cu, Mn, Ni, Pb, Ti, Tl and V concentrations were measured using ICP-mass spectrometry (ICP-MS, PerkinElmer SCIEX, Elan 9000, Norway);

Table 1
Method and input variables of the four developed models.

	Method	Input variables
Model I	BP-ANN	Atmospheric pollutants
Model II	BP-ANN	PM ₁ + Meteorological factors
Model III	SVM	Atmospheric pollutants
Model IV	SVM	PM ₁ + Meteorological factors

the minimum detection limits for most of those elements in solution was 0.01 µg/L. ICP-MS was optimized using solutions of Mg, Rh, In, Ba, Ce, Pb and U, each at a concentration of 10 µg/L and prepared in 2% HNO₃. The internal standard in the element analyses was ¹¹⁵In, at a concentration of 20 µg/L and prepared in 2% HNO₃. At least four blank filters were analyzed simultaneously, and the concentration of each element then corrected by subtracting the average blank concentration. Quality control was ensured by analyzing NIST SRM 1648a (urban particulate matter). The recovery rates of the studied elements were between 86% and 105%.

2.3. Health risk assessments

Health risk assessments are used to estimate the occurrence of adverse health effects in children and adults that result from the direct inhalation of atmospheric particulates (US EPA, 1989, 2009). The risk assessment model mainly involves exposure assessment and risk characterization. In the latter, the reference concentration (RfC) and inhalation unit risk (IUR) are used to characterize the risks posed by elements with noncarcinogenic (i.e., As, Cd, Cr, Mn, Ni, Tl and V) and carcinogenic (i.e., As, Cd, Cr(VI), Ni and Pb) effects, respectively. The calculations of the inhalation exposure concentration (EC), the hazard quotient (HQ) of the noncarcinogenic and carcinogenic risks (CR) are shown in the Supporting Information (SI).

2.4. Simulation models

In this study, the machine learning models BP-ANN and SVM were used to simulate the concentrations of particle-bound heavy metals. Atmospheric pollutant concentrations (SO₂, NO₂, CO, O₃ and PM_{2.5}), meteorological factors (temperature, relative humidity, atmospheric pressure and wind speed) and the PM₁ concentration served as input variables. A total of four models were developed according to the input variables and methods (Table 1). All data were randomly partitioned into two sets, with 75% for the training set and 25% for the test set. The optimal BP-ANN and SVM models were chosen based on higher correlation coefficients and lower errors in the training and test stages.

A BP learning algorithm was proposed by Rumelhart et al., 1986 (Rumelhart et al., 1986). The BP-ANN is constructed on the basis of multiple feed-forward networks composed of input, output and hidden layers (Zhao et al., 2018). The general idea of BP-ANN is to learn a certain number of samples after input and expected output assigned. The input is sent to each neuron in the input layer and after calculation in the hidden layer and output layer, each neuron in the output layer receives the corresponding predicted concentration of one element. If the error between the predicted concentration and the expected output does not meet the accuracy requirement, the error is propagated from the output layer such that the weight and threshold are adjusted until the accuracy requirement is met. In this study, MATLAB R2013a was used to establish the BP-ANN models.

The SVM models were established using MATLAB R2013a and libsvm-3.21. SVM was initially applied to solve classification problems (Cortes and Vapnik, 1995), but following the introduction of

an ϵ -insensitive loss function nonlinear regression estimates became possible. SVM adopts the structural risk minimization principle and has a good generalization ability (Wang and Hu, 2015). In SVM, linear or nonlinear models are used to project input vectors into a high-dimensional feature space such that complex input-output relationships can be identified in a relatively simple manner (Feng et al., 2020). Given the training set $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, the linear regression function established in higher-dimensional space can be defined as shown in Eq. (1):

$$f(x) = w \cdot \phi(x) + b \quad (1)$$

where $f(x)$ is the predicted values, x the input vector in the sample space, $\phi(x)$ the input variables after kernel transformation, w the weight and b the bias parameter, estimated by minimizing the regularized risk function.

The optimization function and constraints are shown in Eq. (2):

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (2)$$

and the constraint conditions in Eq. (3):

$$\text{s.t.} \begin{cases} f(x_i) - y_i \leq \epsilon + \xi_i \\ y_i - f(x_i) \leq \epsilon + \xi_i^* \\ \xi_i \geq 0, \xi_i^* \geq 0, i = 1, 2, \dots, n \end{cases} \quad (3)$$

where $1/2\|w\|^2$ is the regularization term and a measurement of function flatness; ξ_i and ξ_i^* are relaxation variables, with prediction errors indicated by values > 0 , otherwise both are equal to 0; C is the penalty factor used to reduce the fitting error and ϵ is a prescribed parameter. To optimize the functionality of the SVM model, we focused on C and ϵ .

By introducing Lagrange multipliers and exploiting the optimality constraints, the final nonlinear fitting function can be expressed as shown in Eq. (4):

$$f(x) = \sum_{i=1}^m (d_i - y_i) K(x_i, x) + b \quad (4)$$

In general, polynomial function, sigmoidal function and radial basis function are the most commonly used for SVM models (Wang et al., 2015). In this study, the radial basis function was chosen because it is effective and fast in the training process (Zhao et al., 2016), which is defined as shown in Eq. (5):

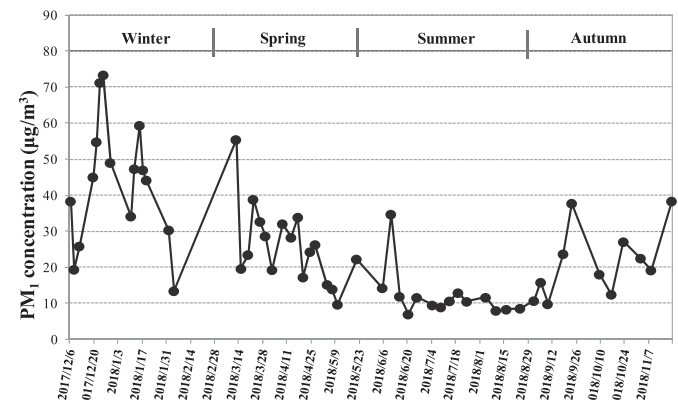


Fig. 1. PM₁ concentrations during the sampling periods.

Table 2
Seasonal variation in the volume-related concentrations of heavy metals in PM₁ (ng/m³).

	Spring	Summer	Autumn	Winter	Annual
As	2.184 ± 1.311	0.807 ± 0.449	1.159 ± 0.699	2.404 ± 1.123	1.706 ± 1.184
Cd	0.374 ± 0.302	0.117 ± 0.076	0.196 ± 0.134	0.632 ± 0.303	0.344 ± 0.305
Co	0.138 ± 0.096	0.044 ± 0.026	0.084 ± 0.053	0.177 ± 0.091	0.115 ± 0.090
Cr	5.936 ± 1.868	3.573 ± 0.816	6.146 ± 1.983	8.733 ± 3.171	6.132 ± 2.798
Cu	9.733 ± 5.764	6.486 ± 3.652	9.709 ± 10.72	25.33 ± 15.03	13.04 ± 12.09
Mn	12.41 ± 7.073	4.607 ± 1.923	10.47 ± 5.881	19.83 ± 9.328	12.07 ± 8.583
Ni	6.224 ± 2.029	3.616 ± 0.892	3.376 ± 1.956	7.756 ± 3.161	5.437 ± 2.799
Pb	15.66 ± 8.681	5.622 ± 2.187	10.36 ± 7.791	26.05 ± 11.32	14.91 ± 11.16
Ti	18.67 ± 13.83	7.596 ± 4.029	14.42 ± 5.370	14.60 ± 5.88	14.06 ± 9.444
Tl	0.159 ± 0.091	0.061 ± 0.039	0.104 ± 0.065	0.275 ± 0.107	0.155 ± 0.114
V	2.133 ± 1.199	1.315 ± 0.746	1.790 ± 1.632	2.323 ± 1.657	1.916 ± 1.362
Zn	45.75 ± 24.69	26.81 ± 14.71	38.78 ± 26.20	95.48 ± 42.93	52.84 ± 38.96

$$K(x_i, x) = \exp \left\{ - \|x - x_i\| / 2\sigma^2 \right\} \quad (5)$$

where, d_i and y_i are the introduced Lagrange multipliers and σ^2 is the width of the Gaussian kernel.

2.5. Evaluation of model performance

The correlation coefficient (R), mean absolute error (MAE), root mean squared error (MSE) and index of agreement (IA) were used to evaluate the performances of the four models (described in the SI). In general, R is used to measure the model's fit performance; MAE and MSE are indicators of the residual errors and are used to evaluate the model's predictive effectiveness; and IA is used to measure the similarity between the modeled and the observed tendencies, with values closer to 1 indicating greater similarity.

3. Results and discussion

3.1. Mass concentrations of PM₁

The temporal trend in the PM₁ concentration from December 2017 to November 2018 in Nanjing is shown in Fig. 1. The seasonal variations in meteorological factors and other air pollutants are summarized in Table S1. The annual average PM₁ concentration was 26.31 µg/m³, below the annual standard limit of 35 µg/m³ set by the National Ambient Air Quality Standard (NAAQS) in China (GB3095-2012) for PM_{2.5} but much higher than the annual guideline value of 10 µg/m³ proposed by the WHO for PM_{2.5}. The average PM₁ concentration decreased from a high in winter of 43.56 µg/m³ (range: 13.48–73.40 µg/m³) to 25.98 µg/m³ in spring (9.72–55.46 µg/m³), 21.43 µg/m³ in autumn (9.84–38.38 µg/m³) and 12.07 µg/m³ in summer (7.00–34.75 µg/m³). The higher concentration of PM₁ in winter was mainly due to the increased domestic heating and relatively stable meteorological conditions characteristic of this season, including low temperatures and light wind speeds, which enhance the accumulation of air pollutants (Zhang et al., 2015). By contrast, the lower PM₁ concentrations during summer can be attributed to the characteristic high temperatures, abundant rain and relatively strong diffusion capacity (Li et al., 2018). The concentrations of SO₂, NO₂, CO and PM_{2.5} were also highest in winter and lowest in summer, with the exception of O₃, the concentration of which was higher in summer and spring and lower in winter.

3.2. Heavy metal concentrations

The volume- and mass-based concentrations of heavy metals during the four seasons are presented in Table 2 and Table S2, respectively. Zn, Pb and Ti were abundant in PM₁ whereas the

concentrations of Cd, Tl and Co were low. For most elements described by volume-related concentrations, the seasonal decrease followed the order winter > spring > autumn > summer, except Ti, Cr and Ni. The Ti concentration was highest in spring (Table 2) whereas the Cr concentration was higher in autumn than in spring, and the Ni concentration was slightly higher in summer than in autumn. Although metals expressed as volume-related concentrations are commonly used to assess their pollution status, their expression as a proportion of PM mass also provides information of their toxicity. The seasonal variation of mass-related concentrations of heavy metals were mainly related with their different pollution sources. Among the metals described by mass-related concentrations, As, Cd, Cu, Pb and Tl concentrations were highest in winter whereas the concentrations of As, Cd, Co, Mn, Ni, Pb and Tl were lowest in summer (Table S2).

As shown in Table S3, for heavy metals in PM₁, the concentrations recorded in Nanjing differed from those determined in other cities because of differences in background values and pollution levels (Caggiano et al., 2010; Huang et al., 2016; Li et al., 2019; Liu et al., 2014; Mainka and Zajusz-Zubek, 2019; Onat et al., 2013; Perrone et al., 2013; Talbi et al., 2018; Trippetta et al., 2016; Zajusz-Zubek et al., 2017). In general, the concentrations of Co and V determined in this study were lower than in all other studies, and the concentrations of As and Cd were lower in Nanjing than in most other cities. The concentrations of Cr, Cu, Mn, Ni, Pb, Ti, Tl and Zn were in the middle range of the previously published values.

Anthropogenic influences were distinguished from the natural background content of metals by calculating an enrichment factor (EF) based on the normalization of a given metal with respect to a reference element (see the SI). As shown in Fig. S1, the average EF values of Co, Mn and V during the four seasons were < 10, indicative of a minimal enrichment of these metals and their having originated mainly from crustal sources (Tang and Han, 2017). Cr, Ni and Tl were moderately enriched (10 < EF < 100), and As, Cd, Pb and Zn greatly enriched (EF > 100) during all four seasons. Cu was moderately enriched in spring and autumn (10 < EF < 100) but greatly enriched (EF > 100) in summer and winter. The EF values of nearly all of the studied metals were highest in winter and lowest in summer or autumn. The exception was Cu, whose EF was lowest in spring. These anthropogenic elements in the study area mainly derived from industrial emissions, coal combustion and traffic, etc (Li et al., 2017).

3.3. Health risk posed by toxic metals in PM₁

The health risk caused by direct exposure to airborne metals via inhalation is shown in Table S4. The EC values of Mn, Ni and Pb were higher than those of the other tested metals. For noncarcinogens, Mn and Ni had the highest HQ values in both children and adults.

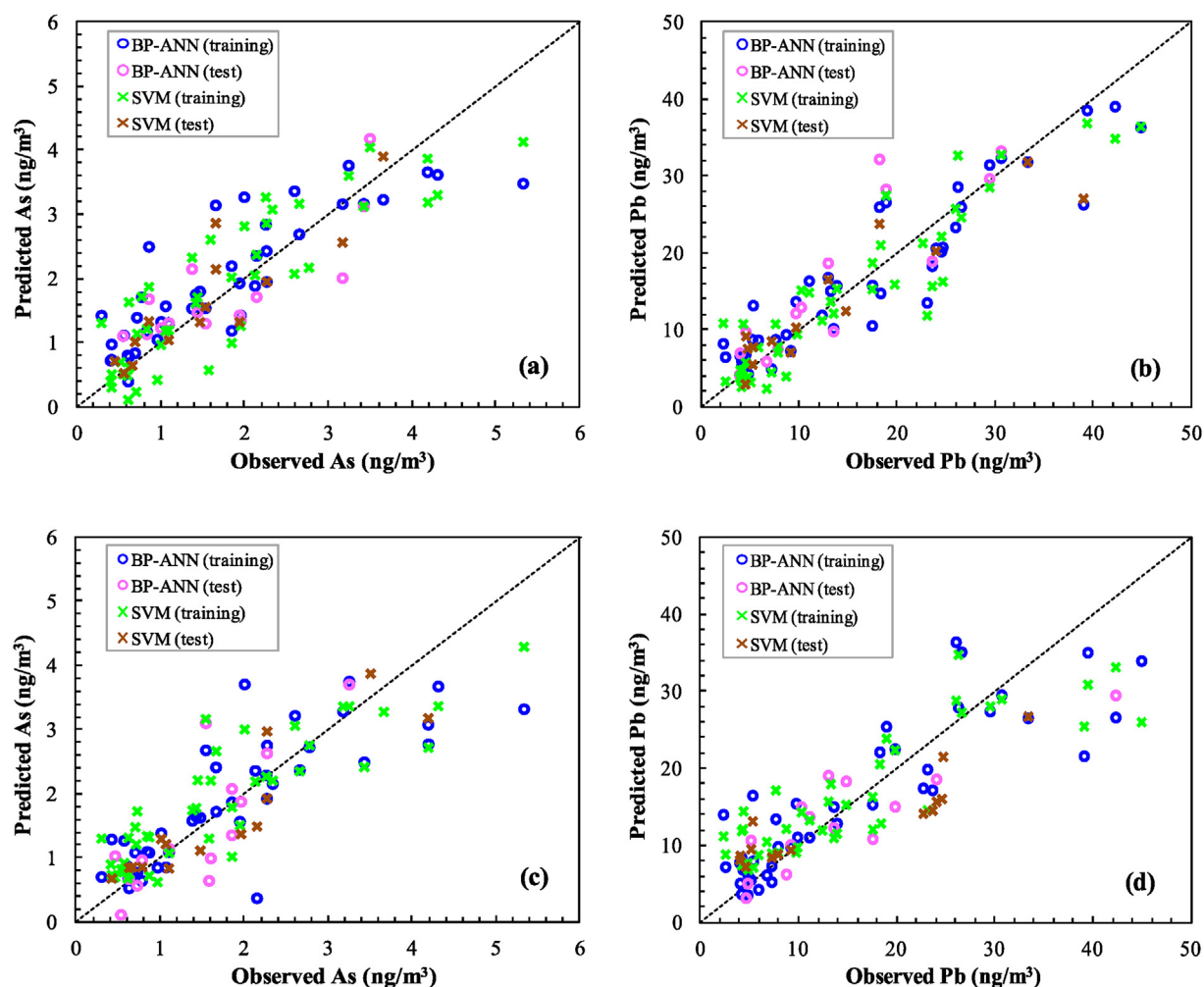


Fig. 2. Predicted vs. observed concentrations of As (a, c) and Pb (b, d) at the training and test stages as simulated by the BP-ANN and SVM models. Atmospheric As and Pb levels were simulated in (a) and (b) both using atmospheric pollutants, as well as in (c) and (d) both using meteorological factors and PM_1 concentrations, respectively.

The HQ values resulting from inhaled As, Cd, Co, Cr, Mn, Ni, Tl and V in PM_1 were all below the safe limit (1) for these two sub-populations. The hazard index (HI) was 0.419, which was below the safe limit (1) and thus indicated no obvious accumulative noncarcinogenic risks. Among the carcinogens, Cr had the highest risk whereas the carcinogenic risks for the inhalation of As, Cd, Co, Ni, and Pb were all less than the precautionary value set for children and adults (10^{-4}). The combined carcinogenic risk was 3.61×10^{-6} for children and 1.45×10^{-5} for adults, which in both cases was lower than the precautionary value.

However, there were several uncertainties in the assessment models, exposure metal toxicity data and exposure parameters of the populations and thus in the health risk results. In addition, other pathways of atmospheric toxic metal exposure were not considered, such as the inadvertent ingestion of atmospheric particulates and dermal absorption via exposed skin. Finally, other potentially toxic heavy metals, such as Hg, Sb, Cu and Zn, were not considered in this study. Nonetheless, previous studies have demonstrated the validity of this health risk assessment method as an effective tool for evaluating the adverse health effects caused by exposure to airborne heavy metals (Sun et al., 2014; Hsu et al., 2016; Caggiano et al., 2019).

3.4. Correlations among atmospheric pollutants, meteorological factors and metal concentrations

The Spearman's correlation coefficients of atmospheric pollutants, meteorological factors and the volume-related heavy metal concentrations in PM_1 are summarized in Tables S5 and S6. Nearly all of the studied metals correlated significantly and positively with each other (except Ti with Ni and V), indicating their common origin (Table S5). The PM_1 concentration correlated positively with almost all of the heavy metals, as evidenced by correlation coefficients ranging from 0.467 for V to 0.885 for Tl. For most of the tested metals, the correlations with temperature and wind speed were significantly negative, and the correlations with atmospheric pressure significantly positive (except V with temperature, Ni and V with wind speed, and V with atmospheric pressure). Only Ti correlated significantly with relative humidity. All of the metals correlated significantly with atmospheric pollutants (SO_2 , NO_2 , CO and $PM_{2.5}$). The latter finding reflects the fact that heavy metals in PM_1 and atmospheric pollutants derive from similar sources, including coal combustion, vehicle exhaust, biomass burning, and road dust (Caggiano et al., 2019; Tao et al., 2014). SO_2 and NO_2 are the gaseous precursors of secondary inorganic components of PM_1 , and thus play important roles in determining its concentration as well as that of PM_1 -bound heavy metals (Li et al., 2019). The

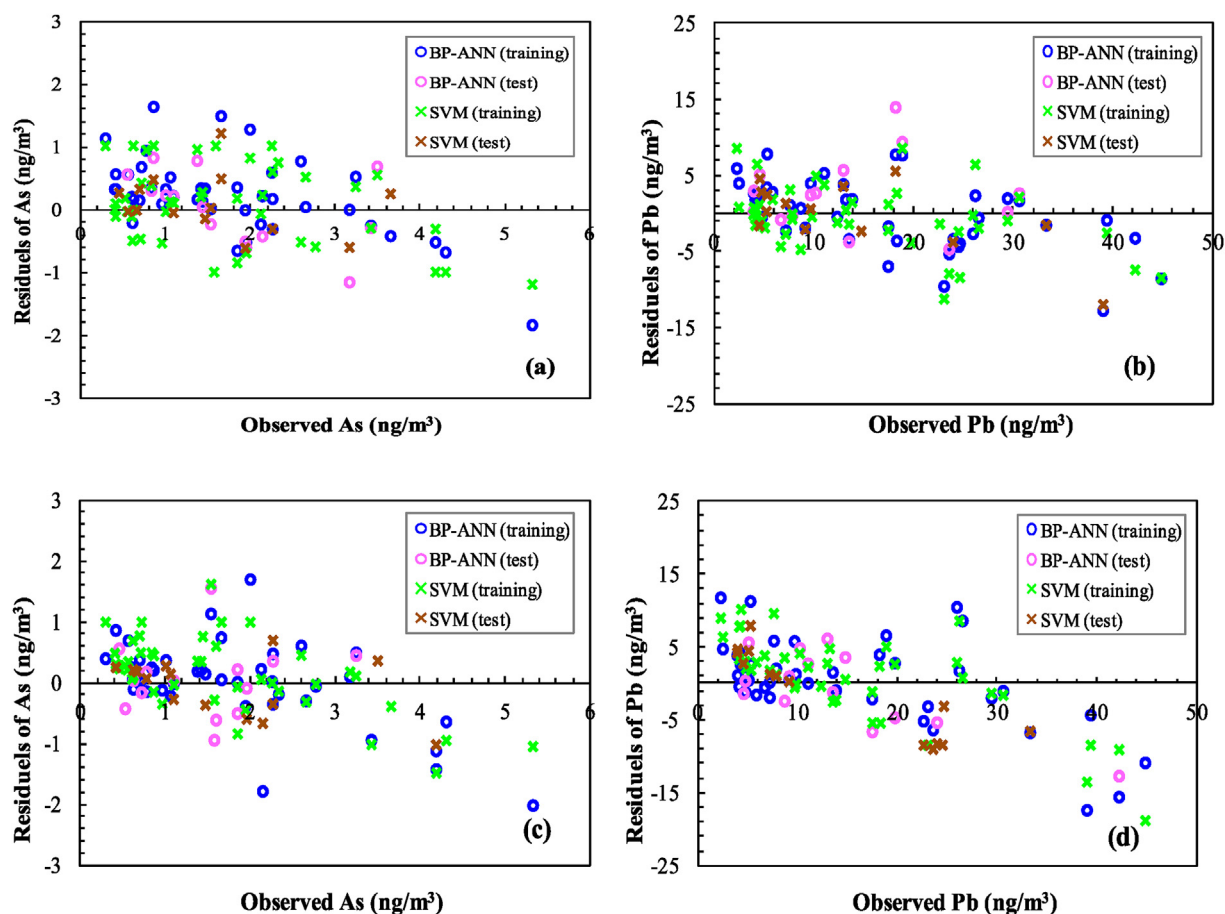


Fig. 3. Residuals plot of As (a, c) and Pb (b, d) at the training and test stages simulated by the BP-ANN and SVM models. Atmospheric As and Pb levels were simulated in (a) and (b) both using atmospheric pollutants, as well as in (c) and (d) both using meteorological factors and PM_1 concentrations, respectively.

significant negative correlations of the studied metals with O_3 (except Ti and V, which had no significant correlation with O_3) can be attributed to its being a ground-surface-derived secondary pollutant whose formation involves complex photochemical reactions that differ from those of primary atmospheric pollutants (Li and Cocker, 2018).

3.5. Simulation results

The volume-related concentrations of heavy metals in PM_1 were simulated using BP-ANN and SVM models (Table 1). The results, expressed as R, MAE and RMSE, are summarized in Tables S7–S10. The predicted versus the observed values and the corresponding residuals were plotted for As and Pb, two international routine monitoring indicators (Figs. 2 and 3, respectively).

The training results determine the model performances of the networks. For most elements, a higher training or test R value was generally linked to lower MAE and RMSE values and to a higher IA value. The training R and test R values of As, Cr, Cu, Mn, Pb, Tl and Zn were all >0.8 in the BP-ANN models (models I and II), whereas the training R values of As, Co, Pb, Tl and Zn were all >0.8 in the SVM models (models III and IV). The training R and test R values of Co, Ti, and V were <0.8 in the BP-ANN models, and those of Ni, Ti and V <0.8 in the SVM models. In general, all four models performed better for Pb, Tl and Zn, based on training R and test R values consistently >0.85 , lower errors and higher IA values, whereas the performances of all four models for Ti and V were relatively poor, as the training R and test R values were <0.8 . According to the EF

(Fig. S1), Pb and Zn were greatly enriched, Tl was moderately enriched, and V and Ti originated mainly from crustal sources. These results indicated strong relationships between the input variables and the PM_1 -associated heavy metals that derived from anthropogenic activities, such as industrial emissions and traffic activities, but weaker relationships with metals that derived from natural processes. In our previous study, better prediction results for Ni, Al, V, Cd and As in size-fractionated PM were obtained in both the BP-ANN and the SVM models using meteorological factors and the PM concentrations as input parameters (Leng et al., 2017). In another study from our group, better prediction results for Al, Fe, Mn, Ni, and Ti were obtained in the SVM models when meteorological factors, $PM_{2.5}$ concentrations and the magnetic properties of $PM_{2.5}$ were the input parameters (Li et al., 2017). Atmospheric V, Mn and Ti derive from natural process whereas Ni, Cd, As are mainly from anthropogenic sources. Thus, in our two previous studies the different sources of the heavy metals likely accounted for the complexity of the prediction results. However, according to the present study, the modeling effects of particle-bound metals in nonlinear models also closely depend on PM size.

A comparison of the BP-ANN and SVM models showed that the training R values of Cr, Cu, Mn, Ni, Pb, Tl and Zn were higher in the BP-ANN models than in the SVM models when atmospheric pollutants were used as input variables. In addition, the training R values of Cd, Cr, Cu, Mn, Ni, Pb, Ti and Tl were higher in the BP-ANN models than in the SVM models when the input variables were PM_1 and meteorological factors. In general, BP-ANN performed better than SVM for Cr, Cu, Mn, Ni, Pb and Tl, and SVM better than BP-ANN

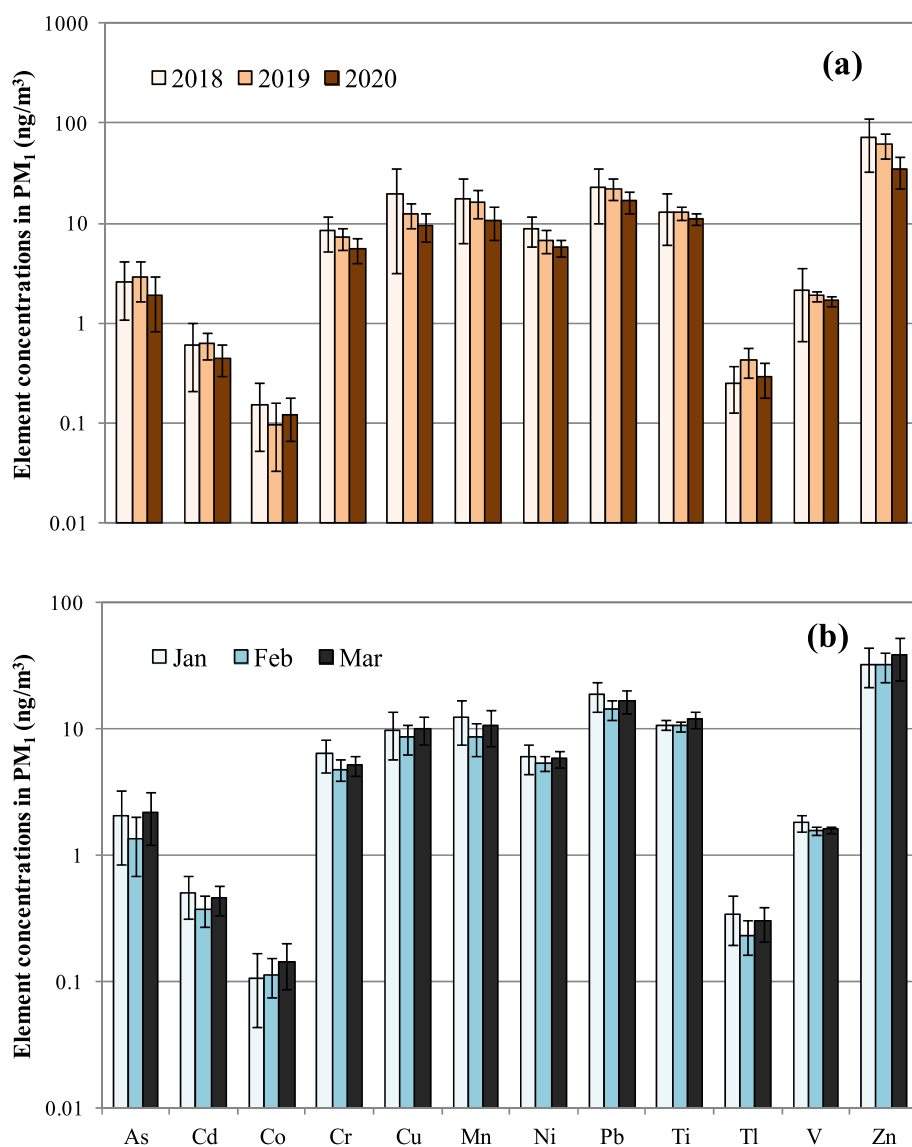


Fig. 4. (a) Comparison of the mean predicted element concentrations in PM_{10} from January to March 2019 and 2020, and the observed elemental concentrations in 2018. (b) Comparison of the mean predicted element concentrations in PM_{10} in January, February and March 2020 (b).

for As, Cd, Co and V. The advantages of ANN in non-linear systems, especially for difficult to construct theoretical models, are well-known (Gardner and Dorling, 1998), such that this approach has been frequently used as a nonlinear tool, including in recent air quality forecasting studies (Cabaneros et al., 2019). However, ANN has a strong dependence on historical data and the results are likely to be poor if few samples are available (Feng et al., 2011). SVM avoids the overfitting and dimension disaster problems of BP-ANN and thus has a greater potential to regress the input-output relationship during the training phase, resulting in a good performance for new input data (Yeganeh et al., 2012). Although in our previous studies the SVM models performed slightly better than the BP-ANN models in the simulation of many heavy metals in $PM_{2.5}$ (Leng et al., 2017), in the present study there were no obvious differences in their simulation performances.

A comparison of the models' performances using different input variables showed that for As, Cd, Cu, Mn, Ni, Pb and Zn the training R values were higher when the input variables were atmospheric pollutants rather than PM_{10} and meteorological factors. This was the

case for both the BP-ANN and the SVM models. By contrast, for Cr, Ti, and V, the training R values were higher when the input variables were PM_{10} and meteorological factors rather than atmospheric pollutants, again for both the BP-ANN and the SVM models. In either case, there was no dependency on the linear correlation between a target element and the input variables, because for a target element high correlation coefficients with the input variables did not always correspond to high training R values (Table S6), thus providing further evidence of the strong nonlinear relationships between the metal concentrations and the input variables. The EF values of As, Cd, Cu, Ni, Pb and Zn were much higher than those of Cr and V, indicating that atmospheric pollutants are good indicators in predictions of the anthropogenic elements in PM_{10} using nonlinear models.

3.6. Prediction of metal concentrations in PM_{10} from Nanjing during 2019–2020

Following the successful establishment of the simulation

models of heavy metals in PM₁, the metal concentrations in PM₁ from Nanjing between January 1, 2019 and March 31, 2020 were predicted using the fully trained SVM model and, as input variables, the daily concentrations of five atmospheric pollutants (SO₂, NO₂, CO, O₃ and PM_{2.5}) as recorded by the Nanjing Environmental Protection Bureau. The results are shown in Tables S11 and S12 and in Fig. 4.

The highest predicted concentration in PM₁ was that of Zn, followed by Pb, Mn and Ti, and the lowest concentrations were those of Co, Tl, and Cd. These findings are consistent with the observed results for 2018. With the exceptions of Cr, Cu, Ti, V and Zn, the mean annual heavy metal concentrations predicted in 2019 were higher than the observed concentrations in 2018. This was the case for As, Cd, Co, Mn, Ni, Pb and Tl (Table S11). Seasonal decreases in the concentrations of Cd, Cr, Cu, Ni, Pb, Tl, V and Zn were predicted along the order: winter > spring > autumn > summer. For these elements, pollution was heaviest in January or December and lightest in August or July (Table S12). By contrast, the seasonal decreases in the concentrations of As and Mn followed the pattern spring > winter > autumn > summer, and those of Co and Ti the pattern spring > autumn > summer > winter.

The COVID-19 pandemic struck China at the beginning of 2020 (Flahault, 2020). To limit the spread of the disease, China implemented a series of nationwide countermeasures that went into effect from the end of January 2020 until the beginning of March 2020, including traffic control, the shutdown of businesses and quarantine at home (Okoyere et al., 2020). As shown in Fig. S2, except O₃, the mean concentrations of other atmospheric pollutants monitored by the Nanjing Environmental Protection Bureau from January to March 2020 were all lower than those measured from January to March of 2018 and 2019. According to the fully trained SVM models, the mean concentrations of most elements, except Co and Tl, predicted for the period during the COVID-19 pandemic were also lower than either the predicted concentrations from January to March 2019 or the observed concentrations from January to March 2018 (Fig. 4(a)). Furthermore, a comparison of the mean element concentrations for the first three months of 2020 showed that the concentrations of nearly all the studied elements except Co were lowest in February and then increased to some extent in March (Fig. 4(b)). This result is consistent with the implementation period of the nationwide countermeasures, which had significant effects on the emissions of air pollutants.

4. Conclusions

In this study, the heavy metals in PM₁ from Nanjing, China were investigated. During 2018, the annual average PM₁ concentration was 26.31 µg/m³ and most element concentrations followed a seasonal pattern: winter > spring > autumn > summer. As, Cd, Pb and Zn were greatly enriched, whereas Co, Mn and V originated mainly from crustal sources. There were no accumulative noncarcinogenic or carcinogenic risks resulting from inhalation of the toxic metals in PM₁. BP-ANN and SVM models of the heavy metals in PM₁ performed well for Pb, Tl and Zn but relatively poorly for Ti and V. The concentrations of many metals were better simulated when the input variables were atmospheric pollutants rather than PM₁ and meteorological factors, both for the BP-ANN and the SVM models. PM₁-bound heavy metal concentrations in 2019 were predicted by the fully trained models and their concentrations were shown to be higher in winter and lower in summer. During the COVID-19 pandemic from January to March 2020, the predicted concentrations of most metals followed a trend consistent with the implementation of emergency countermeasures in China. These results demonstrated the validity of the simulation models to predict heavy metals in PM₁.

Authorship Contribution Statement

Huiming Li: Writing - original draft, Funding acquisition. Qianying Dai: Writing - original draft. Meng Yang: Methodology. Fengying Li: Data curation. Xuemei Liu: Formal analysis. Mengfan Zhou: Investigation. Xin Qian: Writing - review & editing, Supervision, Conceptualization, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was financed by the National Natural Science Foundation of China (grant no. 41771533), the Natural Science Foundation of Jiangsu Province, China (grant no. BK20171339) and the Open Fund of State Key Laboratory of Pollution Control and Resources Reuse (grant no. PCRRF19026).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.chemosphere.2020.127571>.

References

- Cabaneros, S.M., Calautit, J.K., Hughes, B.R., 2019. A review of artificial neural network models for ambient air pollution prediction. *Environ. Model. Software* 119, 285–304.
- Caggiano, R., Macchiato, M., Trippetta, S., 2010. Levels, chemical composition and sources of fine aerosol particles (PM₁) in an area of the Mediterranean basin. *Sci. Total Environ.* 408, 884–889.
- Caggiano, R., Sabia, S., Speranza, A., 2019. Trace elements and human health risks assessment of finer aerosol atmospheric particles (PM₁). *Environ. Sci. Pollut. Res.* 26, 36423–36433.
- Cortes, C., Vapnik, V., 1995. Support-vector networks. *Mach. Learn.* 20, 273–297.
- Fang, T., Zeng, L.H., Gao, D., Verma, V., Stefaniak, A.B., Weber, R.J., 2017. Ambient size distributions and lung deposition of aerosol dithiothreitol-measured oxidative potential: contrast between soluble and insoluble particles. *Environ. Sci. Technol.* 51, 6802–6811.
- Feng, S., Gao, D., Liao, F., Zhou, F., Wang, X., 2016. The health effects of ambient PM_{2.5} and potential mechanisms. *Ecotoxicol. Environ. Saf.* 128, 67–74.
- Feng, Y., Zhang, W.F., Sun, D.Z., Zhang, L.Q., 2011. Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and support vector machine data classification. *Atmos. Environ.* 45, 1979–1985.
- Feng, Z.K., Niu, W.J., Tang, Z.Y., Jiang, Z.Q., Xu, Y., Liu, Y., Zhang, H.R., 2020. Monthly runoff time series prediction by variational mode decomposition and support vector machine based on quantum-behaved particle swarm optimization. *J. Hydrol.* 583, 124627.
- Flahault, A., 2020. COVID-19 cacophony: is there any orchestra conductor? *Lancet* 395 (10229), 1037.
- Fuertes, E., Sunyer, S., Gehring, U., Porta, D., Forastiere, F., Cesaroni, G., Vrijheid, M., Guxens, M., Annesi-Maesano, I., Slama, R., Maier, D., Kogevinas, M., Bousquet, J., Chatzi, L., Lertxundi, A., Basterrechea, M., Esplugues, A., Ferrero, A., Wright, J., Mason, D., McEachan, R., Garcia-Aymerich, J., Jacquemin, B., 2020. Associations between air pollution and pediatric eczema, rhinoconjunctivitis and asthma: a meta-analysis of European birth cohorts. *Environ. Int.* 136, 105474.
- Gardner, M.W., Dorling, S.R., 1998. Artificial neural networks (the multilayer perceptron)-a review of applications in the atmospheric sciences. *Atmos. Environ.* 32, 2627–2636.
- Hsu, C.Y., Chiang, H.C., Lin, S.L., Chen, M.J., Lin, T.Y., Chen, Y.C., 2016. Elemental characterization and source apportionment of PM₁₀ and PM_{2.5} in the western coastal area of central Taiwan. *Sci. Total Environ.* 541, 1139–1150.
- Huang, L.K., Wang, W., Wang, K., Hong, L.J., Wang, J.Y., 2016. Pollution characteristics of metal elements in PM₁ during heating period in Harbin. *J. Natural. Sci. Heilongjiang Univ.* 33, 808–812 (in Chinese).
- Kelly, F.J., Fussell, J.C., 2012. Size, source and chemical composition as determinants of toxicity attributable to ambient particulate matter. *Atmos. Environ.* 60, 504–526.
- Leng, X.Z., Wang, J.H., Ji, H.B., Wang, Q.G., Li, H.M., Qian, X., Li, F.Y., Yang, M., 2017. Prediction of size-fractionated airborne particle-bound metals using MLR, BP-ANN and SVM analyses. *Chemosphere* 180, 513–522.

- Li, H.M., Wang, J.H., Wang, Q.G., Tian, C.H., Qian, X., Leng, X.Z., 2017. Magnetic properties as a proxy for predicting fine-particle-bound heavy metals in a support vector machine approach. *Environ. Sci. Technol.* 51, 6927–6935.
- Li, K.W., Chen, L.H., White, S.J., Zheng, X.J., Lv, B., Lin, C., Bao, Z.E., Wu, X.C., Gao, X., Ying, F., Shen, J.D., Azzi, M., Cen, K.F., 2018. Chemical characteristics and sources of PM₁ during the 2016 summer in Hangzhou. *Environ. Pollut.* 232, 42–54.
- Li, N., Han, W.Z., Wei, X., Shen, M.N., Sun, S., 2019. Chemical characteristics and human health assessment of PM₁ during the Chinese spring festival in Changchun, northeast China. *Atmos. Pollut. Res.* 10, 1823–1831.
- Li, R., Cui, L.L., Liang, J.H., Zhao, Y.L., Zhang, Z.Y., Fu, H.B., 2020. Estimating historical SO₂ level across the whole China during 1973–2014 using random forest model. *Chemosphere* 247, 125839.
- Li, W.H., Cocker III, D.R., 2018. Secondary organic aerosol and ozone formation from photo-oxidation of unburned diesel fuel in a surrogate atmospheric environment. *Atmos. Environ.* 184, 17–23.
- Liu, L., Hu, H., Li, X., Huang, H.Y., Cai, X.J., Zhang, G.F., 2014. Pollution characteristics and source apportionment of heavy metals in PM₁ at Dongguan, China. *Acta Sci. Circumstantiae* 34, 303–309 (in Chinese).
- Lu, W.Z., Wang, D., 2014. Learning machines: rationale and application in ground-level ozone prediction. *Appl. Soft Comput.* 24, 135–141.
- Luna, A.S., Paredes, M.L.L., de Oliveira, G.C.G., Corrêa, S.M., 2014. Prediction of ozone concentration in tropospheric levels using artificial neural networks and support vector machine at Rio de Janeiro, Brazil. *Atmos. Environ.* 98, 98–104.
- Mainka, A., Zajusz-Zubek, E., 2019. PM₁ in ambient and indoor air—urban and rural areas in the upper silesian region, Poland. *Atmosphere* 10, 662.
- Niu, M.F., Wang, Y.F., Sun, S.L., Li, Y.W., 2016. A novel hybrid decomposition-and-ensemble model based on CEEMD and GWO for short-term PM_{2.5} concentration forecasting. *Atmos. Environ.* 134, 168–180.
- Oberdörster, G., Oberdörster, E., Oberdörster, J., 2005. Nano toxicology: an emerging discipline evolving from studies of ultrafine particles. *Environ. Health Perspect.* 113, 823–839.
- Okyere, M.A., Forson, R., Essel-Gaisey, F., 2020. Positive externalities of an epidemic: the case of the coronavirus (COVID-19) in China. *J. Med. Virol.* 1–4.
- Onat, B., Sahin, U.A., Akyuz, T., 2013. Elemental characterization of PM_{2.5} and PM₁ in dense traffic area in Istanbul, Turkey. *Atmos. Pollut. Res.* 4, 101–105.
- Palani, S., Tkalic, P., Balasubramanian, R., Palanichamy, J., 2011. ANN application for prediction of atmospheric nitrogen deposition to aquatic ecosystems. *Mar. Pollut. Bull.* 1198–1206.
- Perrone, M.R., Becagli, S., Garcia Orza, J.A., Vecchi, R., Dinoi, A., Udisti, R., Cabello, M., 2013. The impact of long-range-transport on PM₁ and PM_{2.5} at a Central Mediterranean site. *Atmos. Environ.* 71, 176–186.
- Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. *Nature* 323, 533–536.
- Sarti, E., Pasti, L., Rossi, M., Ascanelli, M., Pagnoni, A., Trombini, M., Remelli, M., 2015. The composition of PM₁ and PM_{2.5} samples, metals and their water soluble fractions in the Bologna area (Italy). *Atmos. Pollut. Res.* 6, 708–718.
- Sun, Y.Y., Hu, X., Wu, J.C., Lian, H.Z., Chen, Y.J., 2014. Fractionation and health risks of atmospheric particle-bound as and heavy metals in summer and winter. *Sci. Total Environ.* 493, 487–494.
- Taghvaei, H., Amooie, M.A., Hemmati-Sarapardeh, A., Taghvaei, H., 2016. A comprehensive study of phase equilibria in binary mixtures of carbon dioxide plus alcohols: application of a hybrid intelligent model (CSA-LSSVM). *J. Mol. Liq.* 224, 745–756.
- Talbi, A., Kerchich, Y., Kerbach, R., Boughedaoui, M., 2018. Assessment of annual air pollution levels with PM₁, PM_{2.5}, PM₁₀ and associated heavy metals in Algiers, Algeria. *Environ. Pollut.* 232, 252–263.
- Tang, Y., Han, G.L., 2017. Characteristics of major elements and heavy metals in atmospheric dust in Beijing, China. *J. Geochem. Explor.* 176, 114–119.
- Tao, J., Gao, J., Zhang, L., Zhang, R., Che, H., Zhang, Z., Lin, Z., Jing, J., Gao, J., Hsu, S.C., 2014. PM_{2.5} pollution in a megacity of southwest China: source apportionment and implication. *Atmos. Chem. Phys.* 14, 8679–8699.
- Trippetta, S., Sabia, S., Caggiano, R., 2016. Fine aerosol particles (PM₁): natural and anthropogenic contributions and health risk assessment. *Air. Qual. Atmos. Health.* 9, 621–629.
- US EPA (U.S. Environmental Protection Agency), 1989. Risk Assessment Guidance for Super Fund Volume I Human Health Evaluation Manual (Part a). EPA/540/1-89/002 December. <http://www.epa.gov/swerrims/riskassessment/ragsa/index.htm>.
- US EPA (U.S. Environmental Protection Agency), 2009. Risk Assessment Guidance for Superfund (RAGS), Volume I Human Health Evaluation Manual (Part F, Supplemental Guidance for Inhalation Risk Assessment). EPA-540-R-070e002, OSWER 9285.7-82, January. <http://www.epa.gov/swerrims/riskassessment/ragsf/index.htm>.
- Wang, J.Z., Hu, J.M., 2015. A robust combination approach for short-term wind speed forecasting and analysis - combination of the ARIMA (autoregressive integrated moving average), ELM (extreme learning machine), SVM (support vector machine) and LSSVM (least square SVM) forecasts using a GPR (Gaussian process regression) model. *Energy* 93, 41–56.
- Wang, P., Liu, Y., Qin, Z.D., Zhang, G.S., 2015. A novel hybrid forecasting model for PM₁₀ and SO₂ daily concentrations. *Sci. Total Environ.* 505, 1202–1212.
- Wu, Q.L., Lin, H.X., 2019. A novel optimal-hybrid model for daily air quality index prediction considering air pollutant factors. *Sci. Total Environ.* 683, 808–821.
- Ye, Z.P., Yang, J.Q., Zhong, N., Tu, X., Jia, J.N., Wang, J.D., 2020. Tackling environmental challenges in pollution controls using artificial intelligence: a review. *Sci. Total Environ.* 699, 134279.
- Yeganeh, B., Hewson, M.G., Clifford, S., Tavassoli, A., Knibbs, L.D., Morawska, L., 2018. Estimating the spatiotemporal variation of NO₂ concentration using an adaptive neuro-fuzzy inference system. *Environ. Model. Software* 100, 222–235.
- Yeganeh, B., Motlagh, M.S.P., Rashidi, Y., Kamalan, H., 2012. Prediction of CO concentrations based on a hybrid partial least square and support vector machine model. *Atmos. Environ.* 55, 357–365.
- Zajusz-Zubek, E., Radko, T., Mainka, A., 2017. Fractionation of trace elements and human health risk of submicron particulate matter (PM₁) collected in the surroundings of coking plants. *Environ. Monit. Assess.* 189, 389.
- Zhang, Y.W., Zhang, X.Y., Zhang, Y.M., Shen, X.J., Sun, J.Y., Ma, Q.L., Yu, X.M., Zhu, J.L., Zhang, L., Che, H.C., 2015. Significant concentration changes of chemical components of PM₁ in the Yangtze River Delta area of China and the implications for the formation mechanism of heavy haze—fog pollution. *Sci. Total Environ.* 538, 7–15.
- Zhao, Y.S., Zhang, X.P., Deng, L.Y., Zhang, S.J., 2016. Prediction of viscosity of imidazolium-based ionic liquids using MLR and SVM algorithms. *Comput. Chem. Eng.* 92, 37–42.
- Zhao, Y.X., Zhou, D., Yan, H.L., 2018. An improved retrieval method of atmospheric parameter profiles based on the BP neural network. *Atmos. Res.* 213, 389–397.